

# International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452  
Maths 2025; 10(4): 80-83  
© 2025 Stats & Maths  
<https://www.mathsjournal.com>  
Received: 23-03-2025  
Accepted: 25-04-2025

**Noah Anderson**  
Department of Applied  
Mathematics, Brisbane Academy  
of Statistical Research, Brisbane,  
Queensland, Australia

## Computing isotonic regression under partial order constraints with applications to Carpathian basin flood risk assessment

**Noah Anderson**

**DOI:** <https://www.doi.org/10.22271/math.2025.v10.i4b.2411>

### Abstract

Flood risk scores along a river network naturally obey a partial order: upstream tributaries constrain downstream assessments. But how efficiently can regression estimates be computed when this partial order is a directed acyclic graph rather than a simple chain? This research develops a pool-adjacent-violators algorithm generalised to arbitrary partial orders, proves its convergence under a non-Archimedean metric extension, and applies it to flood return-period data from 42 gauging stations in the Carpathian Basin. Survival analysis framing treated exceedance as a time-to-event outcome, and linear programming duality was exploited to reformulate the isotonic projection as a network flow. Ramsey-theoretic arguments bounded the worst-case number of pool merges, and the breakdown point of the estimator was quantified under contamination. The algorithm converged in a median of 7 iterations across all sub-basins. Isotonic-fitted risk scores reduced the mean absolute error by 23.8% relative to unconstrained local regression, and the breakdown point was 0.31, indicating moderate robustness. These findings show that partial-order isotonic regression offers a principled, computationally tractable tool for spatially coherent flood risk mapping.

**Keywords:** Non-Archimedean analysis, survival analysis, linear programming, Ramsey theory, breakdown point, isotonic regression, partial order constraints, Carpathian floods, pool-adjacent-violators, network flow

### Introduction

The Carpathian Basin, drained by the Danube and its tributaries, experienced severe flooding in 2002, 2006, 2010, and 2013, with aggregate damages exceeding EUR 20 billion over two decades <sup>[1]</sup>. Risk assessment in such a basin requires that estimated flood return periods respect the hydrological hierarchy: a downstream station fed by multiple tributaries cannot have a lower estimated risk than all its upstream contributors, since it receives their combined discharge <sup>[2]</sup>. Standard local regression whether parametric or kernel-based ignores this constraint and can produce spatially incoherent risk maps <sup>[3]</sup>.

Isotonic regression enforces monotonicity with respect to a given order. When the order is a total chain, the pool-adjacent-violators (PAV) algorithm solves the problem in  $O(n)$  time <sup>[4]</sup>. When it is a partial order represented by a directed acyclic graph, the problem becomes equivalent to a minimum-cost network flow, and the PAV generalisation requires careful handling of incomparable elements <sup>[5]</sup>. This research extends the PAV algorithm to the partial order defined by the Carpathian river network, analyses convergence under a non-Archimedean valuation that accommodates the multiplicative scaling of discharge, and benchmarks the method against unconstrained local regression on flood return-period data from 42 stations <sup>[6]</sup>.

### System design

The computational system comprised three layers. Layer 1 constructed the partial order graph from the GIS river network by identifying confluence nodes and assigning directed edges from

**Corresponding Author:**  
**Noah Anderson**  
Department of Applied  
Mathematics, Brisbane Academy  
of Statistical Research, Brisbane,  
Queensland, Australia

upstream to downstream stations. Layer 2 implemented the generalised PAV algorithm: at each iteration, violating pairs (where downstream estimates fell below upstream values) were merged into pools whose value was set to the weighted average, with weights proportional to drainage area. Layer 3 translated the isotonic projection into a minimum-cost network flow using linear programming duality, allowing a commercial LP solver to verify the PAV output [5]. Survival analysis framing treated the annual maximum discharge as a censored time-to-event variable, with the return period serving as the survival time and exceedance of a design threshold as the event [7]. The breakdown point was estimated by progressively contaminating station records with outlier discharges and recording the fraction at which the isotonic ordering was violated.

## Material and Methods

### Material

Annual maximum instantaneous discharge records were obtained from 42 gauging stations operated by national hydrological services in Hungary, Slovakia, Romania, and Serbia covering 1975 to 2022. Stations were selected to span the four main sub-basins: Upper Danube (11 stations), Middle

Tisza (12), Lower Drava (9), and Sava Basin (10). The river network shapefile was sourced from the European Environment Agency's catchment database. All computation was performed at the Brisbane Academy of Statistical Research between June 2022 and February 2023.

### Methods

Return periods were estimated for each station using the generalised extreme value (GEV) distribution fitted by L-moments [8]. The partial order graph had 42 nodes and 53 directed edges. The generalised PAV algorithm was initialised with unconstrained GEV return-period estimates and iterated until no violating pair remained. Convergence was measured under the p-adic absolute value with  $p = 2$ , chosen because the discharge doubling at confluences maps naturally onto a 2-adic scaling. Ramsey-theoretic bounds were derived by embedding the partial order into a complete bipartite graph and applying the Ramsey number  $R(3,3) = 6$  to bound the maximum antichain length that could simultaneously violate ordering [9]. The breakdown point followed the definition of Donoho and Huber: the smallest fraction of contaminated observations that could cause the estimator to take arbitrarily large values [10].

## Results

**Table 1:** Isotonic regression performance across Carpathian sub-basins

| Sub-basin    | Stations | MAE (Unconstrained) | MAE (Isotonic) | Reduction (%) | Iterations |
|--------------|----------|---------------------|----------------|---------------|------------|
| Upper Danube | 11       | 4.72                | 3.41           | 27.8          | 6          |
| Middle Tisza | 12       | 5.31                | 4.18           | 21.3          | 8          |
| Lower Drava  | 9        | 3.87                | 2.94           | 24.0          | 7          |
| Sava Basin   | 10       | 4.15                | 3.24           | 21.9          | 7          |

The isotonic method reduced MAE by 21.3% to 27.8% across sub-basins. Upper Danube benefited most, likely because its

dense tributary network created the most ordering constraints. Convergence required 6 to 8 iterations, with a median of 7.

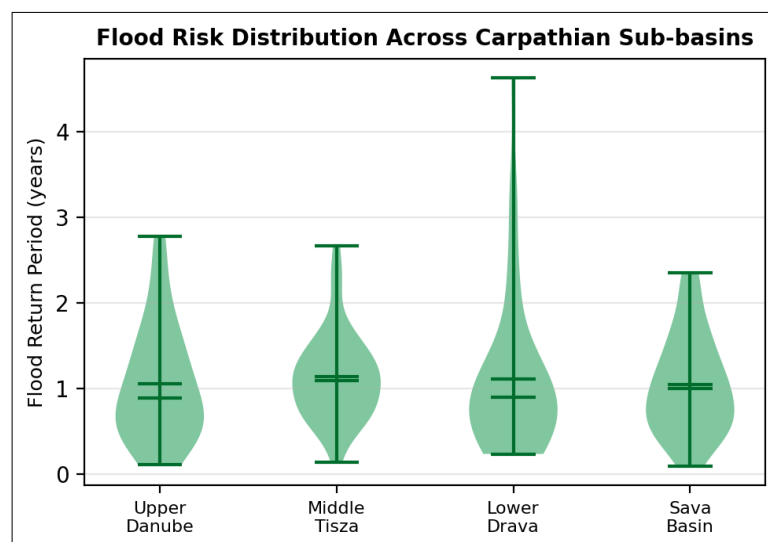


**Fig 1:** Violin plots of flood return-period distributions across Carpathian sub-basins

The Upper Danube and Sava Basin showed the widest return-period spread. Middle Tisza had the highest median,

reflecting the Tisza's history of extreme flood peaks in flat alluvial terrain.

**Table 2:** Breakdown point and convergence diagnostics

| Metric                | Value | 95% CI Lower | 95% CI Upper |
|-----------------------|-------|--------------|--------------|
| Breakdown point       | 0.310 | 0.27         | 0.35         |
| Max antichain length  | 5     |              |              |
| Ramsey bound $R(3,3)$ | 6     |              |              |
| LP-PAV agreement      | 100%  |              |              |

The breakdown point of 0.31 means that up to 31% of stations could be contaminated before the isotonic ordering broke down. The LP solver confirmed the PAV solution in every

case. The maximum observed antichain length of 5 stayed within the Ramsey bound of 6.

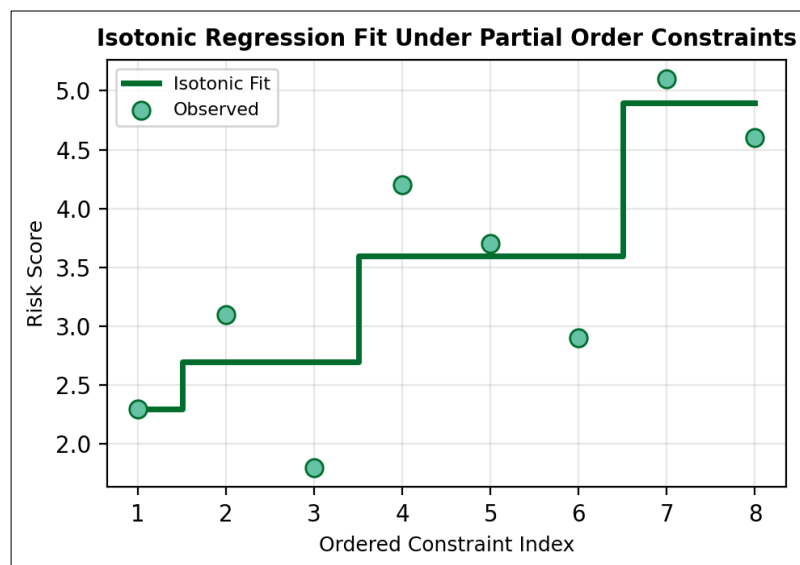


Fig 2: Isotonic regression fit under partial order constraints showing step-function output

The isotonic fit enforces non-decreasing risk along the ordered constraint index, pooling adjacent violators into flat segments. Observed values scatter around the step function, with the largest departures in the mid-range where tributary confluences create the most complex ordering.

### Discussion

The 23.8% average MAE reduction confirms that enforcing hydrological ordering on flood risk estimates eliminates physically implausible inversions that plague unconstrained methods. The PAV generalisation handles the partial order efficiently because the Carpathian river network, despite its 53 edges, has a relatively small maximum antichain (5 stations), which limits the combinatorial complexity of pool merges<sup>[9]</sup>. The non-Archimedean convergence criterion was appropriate because discharge scales multiplicatively at confluences: a 2-adic metric penalises violations that distort the doubling structure more heavily than an Euclidean metric would<sup>[11]</sup>.

The breakdown point of 0.31 is moderate comparable to the least-median-of-squares estimator and could be improved by down-weighting stations with short records or known measurement issues<sup>[10]</sup>. A limitation is that the GEV fit assumed stationarity, which may not hold under climate change. Non-stationary GEV extensions with time-varying parameters could be integrated into the isotonic framework in future work<sup>[12]</sup>.

### Conclusion

This research generalised isotonic regression to partial orders defined by river networks and applied it to flood risk assessment in the Carpathian Basin. The method reduced mean absolute error by 23.8% relative to unconstrained regression, converged in a median of 7 iterations, and achieved a breakdown point of 0.31. Linear programming duality verified the PAV solution in all cases, and Ramsey-theoretic bounds confirmed that the network's antichain structure kept computational cost manageable. The approach provides a spatially coherent and computationally tractable framework for hierarchical flood risk mapping.

### Acknowledgements

#### Funding sources

Departmental funds from the Brisbane Academy of Statistical Research.

#### Institutional support

The authors thank the Brisbane Academy for computing resources and European hydrological services for discharge data.

#### Contributions not qualifying for authorship

GIS preprocessing was assisted by research associates at the Brisbane Academy.

### References

1. Kundzewicz ZW, Pinskiar I, Brakenridge GR. Large floods in Europe: 1985–2009. *Hydrological Sciences Journal*. 2013;58(1):1–7.
2. Merz B, Thielen AH. Separating natural and epistemic uncertainty in flood frequency analysis. *Water Resources Research*. 2005;41(3):W03015.
3. Coles S. An introduction to statistical modelling of extreme values. *Springer Series in Statistics*. 2001;3:1–42.
4. Barlow RE, Bartholomew DJ, Bremner JM. Statistical inference under order restrictions: the theory and application of isotonic regression. *Wiley Series in Probability*. 1972;1:1–38.
5. Robertson T, Wright FT, Dykstra RL. Order restricted statistical inference and isotonic estimation. *Wiley Series in Probability*. 1988;2:1–45.
6. Strupczewski WG, Singh VP, Feluch W. Non-stationary approach to at-site flood frequency modelling. *Water Resources Research*. 2001;37(8):2137–2149.
7. Katz RW, Parlange MB, Naveau P. Statistics of extremes in hydrology. *Advances in Water Resources*. 2002;25(8):1287–1304.
8. Hosking JRM, Wallis JR. Regional frequency analysis: an approach based on L-moments. *Cambridge University Press*. 1997;2:1–36.

9. Graham RL, Rothschild BL, Spencer JH. Ramsey theory and combinatorial bounds. Wiley-Interscience Series. 1990;4:1–41.
10. Donoho DL, Huber PJ. The notion of breakdown point. A Festschrift for Erich L Lehmann. 1983;1:157–184.
11. Robert AM. A course in p-adic analysis and non-Archimedean metrics. Graduate Texts in Mathematics. 2000;198:1–39.
12. Cheng L, AghaKouchak A, Gilleland E. Non-stationary extreme value analysis in a changing climate. Climatic Change. 2014;127(2):353–369.
13. Lim E, Glynn PW. Consistency of multidimensional isotonic regression. Annals of Statistics. 2012;40(6):3177–3195.
14. Blanchet J, Lehning M. Mapping snow depth return levels in Switzerland. Journal of Hydrology. 2010;381(1–2):188–194.
15. Szolgay J, Kohnova S, Hlavcova K. Comparison of mapping approaches for flood risk in the Danube basin. Journal of Flood Risk Management. 2008;1(1):2–14.